

<b>REPORT DOCUMENTATION PAGE</b>			Form Approved OMB NO. 0704-0188	
Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188,) Washington, DC 20503.				
1. AGENCY USE ONLY ( Leave Blank)		2. REPORT DATE December 1, 2011		3. REPORT TYPE AND DATES COVERED Final Report, 4/1/08 – 11/30/11
4. TITLE AND SUBTITLE Graphical, Optimization, and Learning Methods for Fusion and Exploitation in Sensing and Surveillance Systems			5. FUNDING NUMBERS FA9550-08-1-0180	
6. AUTHOR(S) Alan S. Willsky				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Laboratory for Information and Decision Systems Massachusetts Institute of Technology Cambridge, MA 02139			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  Air Force Office of Scientific Research 875 North Randolph Street Arlington, VA 22203			10. SPONSORING / MONITORING AGENCY REPORT NUMBER  AFRL-OSR-VA-TR-2012-0250	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
12 a. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for public release; distribution unlimited.			12 b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This report summarizes our accomplishments under this grant. Our objective is to carry out fundamental research in several interrelated areas: (a) development and use of graphical and hierarchical representations for complex phenomena and for the construction of scalable algorithms for the fusion of heterogeneous sources of information; (b) development of first principles methods for constructing statistical models for the variability of shapes and configurations of objects of interest for statistically optimal shape estimation and object recognition; and (c) development of new adaptive learning and optimization algorithms for analysis of complex, multimodal data for the linking and fusing disparate sources of information, for the characterization of features in complex data and imagery, and for sensor resource management. Our research blends methods from statistics and probabilistic modeling, signal and image processing, optimization, mathematical physics, graphical models, and machine learning theory, yielding new approaches to challenging problems in sensing and surveillance. Moreover, each aspect of our research is directly relevant to Air Force missions. In all of these areas we have contacts and interactions with AFRL staff and with industry involved in Air Force programs.				
14. SUBJECT TERMS Statistical signal and image processing; multiresolution methods, sensor and information fusion; shape modeling, estimation, and recognition; ATR; situational awareness			15. NUMBER OF PAGES	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION ON THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT  UL	

LABORATORY FOR INFORMATION AND DECISION SYSTEMS

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

*Final Technical Report for*

Grant FA9559-08-1-0180

**GRAPHICAL, OPTIMIZATION, AND LEARNING METHODS FOR  
FUSION AND EXPLOITATION IN SENSING AND SURVEILLANCE  
SYSTEMS**

**PRINCIPAL INVESTIGATOR: PROF. ALAN S. WILLSKY**

*Submitted to:* Dr. Jon Sjogren  
AFOSR/NE  
Physics and Electronics  
875 North Randolph Street  
Suite 325, Room 3112  
Arlington, VA 22203

December 1, 2011

## **I. Summary: Objectives and Status of Effort**

In this report we summarize our accomplishments under Grant FA9559-08-1-0180. The objective of this research program is to carry out fundamental research in several interrelated areas: (a) development and use of graphical and hierarchical representations for complex phenomena and for the construction of scalable algorithms for the fusion of heterogeneous sources of information; (b) development of first principles methods for constructing statistical models for the variability of shapes and configurations of objects of interest for statistically optimal shape estimation and object recognition; and (c) development of new adaptive learning and optimization algorithms for analysis of complex, multimodal data for the linking and fusing disparate sources of information, for the characterization of features in complex data and imagery, and for sensor resource management. Our research blends methods from statistics and probabilistic modeling, signal and image processing, optimization, mathematical physics, graphical models, and machine learning theory, yielding new approaches to challenging problems in sensing and surveillance. Moreover, each aspect of our research is directly relevant to Air Force missions. In all of these areas we have contacts and interactions with AFRL staff and with industry involved in Air Force programs.

The principal investigator for this effort is Professor Alan S. Willsky. Prof. Willsky is assisted in the conduct of this research by Dr. John Fisher, principal research scientist in Prof. Willsky's group and by several graduate research assistants as well as additional thesis students not requiring stipend or tuition support from this grant. In the next section we briefly describe our recent research efforts; in Section III we indicate the individuals involved in this effort; in Section IV we list the publications supported by this effort; and in Section V we discuss several other topics including honors received by researchers involved in this project, transitions, and plans for future transitions.

## II. Accomplishments

In this section we briefly describe our research under this grant. We limit ourselves here to a succinct summary and refer to the publications listed at the end of this report for detailed developments. However, we do note here that our work continues to have significant impact, both in terms of DoD-related activities and transitions in progress (Section V) and in terms of recognition from the research community.

### 2.1 Graphical and Hierarchical Models and Scalable Fusion

This component of our research, which has been described in detail in a number of papers and reports [1, 4-5, 7-15, 17, 20, 22, 23, 26-28, 35, 37-38, 40-41, 45, 49-58, 63-81]. The overall objective of this portion of our research is the development of methods for constructing stochastic models for phenomena that vary over space, time, and hierarchy and that possess structure which can be exploited to construct efficient and scaleable algorithms for statistical inference.

- a) We have had a series of successes building on a new approach to inference in Gaussian graphical models that builds on and moves well beyond our previous work on so-called *walk-sum analysis* for inference in Gaussian graphical models. Walk-sum analysis represents an expansion of the set of information made available to a node through successive message passing throughout a graphical model (so that messages engage in “walks” throughout the network during which they are modified at each node, so that information is accumulated in the process). Using this interpretation, we have a precise characterization of the gap between what Belief Propagation computes for error variances in Gaussian models and what the exact computation should produce. This interpretation leads to the tightest known sufficient conditions for BP convergence as well as to a deep understanding of when BP fails. Moreover, this walk-sum analysis has provided the basis for the solution of a long-standing open problem, namely the development of easily checked conditions for the convergence of our previously developed Embedded Trees algorithm. In addition, this work also provides the basis for an adaptive method for choosing which updates should be considered at each stage in the iteration, where the criterion used measures the incremental value-added of each option. Most recently we have taken a much more thorough examination of walks in a graph and in particular on the walk-sums that are not captured by BP. Using the idea of self-avoiding walks we have discovered a representation that makes use of this concept, together with the concept of *cycle bases* from algebraic graph theory, to show how, in principle, exact computation of the variance at a particular node can be computed. The complexity of this computation is closely related to the structure of a graph’s cycle basis and, more specifically to the size of so-called *feedback vertex sets*, i.e., sets of nodes that, when removed from the graph, break all cycles. More importantly, this new insight opens the way to answer a number of important questions, such as (i) developing approximations of increasing quality (but with increasing computational cost) based on incorporating larger and larger subsets of the feedback vertex sets; (ii) efficient sampling from graphical models; and (iii) investigating how computations can be done simultaneously at *all* nodes, something that requires both “header bits” on BP-like messages indicating what nodes each message

has visited as well as memory at each node to remember some of the messages it has received previously. We believe that this investigation will continue to yield very new methods for high-performance inference and especially for distributed fusion algorithms. Experimental results show that in addition to the theoretical guarantees of this method, the approach yields remarkably good results including in essentially all cases in which BP fails to converge.

- b) We have made considerable progress on developing new classes of multiresolution and hierarchical graphical models. For Gaussian processes (or for situations in which we focus on second-order statistics), we have developed a new approach to modeling that represents phenomena at multiple resolutions, with tree-structured statistical relationships *between* scales but with the statistics within each scale, when *conditioned* on other scales, having only local and sparse *correlation structure*. Models of this type yield very efficient algorithms, alternating between rapid tree-structured iterations between scales and local FIR filtering within each scale. We have also adapted ideas from maximum entropy modeling (see paragraph to follow), an approach that in its usual form aims to yield sparse graphical structures, which corresponds to sparse inverse covariance matrices. In our case, we want that sparsity in the portion of the inverse covariance corresponding to the inter-scale behavior, but sparsity in the portion of the *covariance* corresponding to *intra-scale* statistics (conditioned on other scales). We have now demonstrated the power of this method, explained its connections to a generalized notion of ARMA modeling, and written several papers on this approach. In addition, we have made considerable progress in developing analogous methods for discrete-valued processes (and hybrid processes involving both discrete and continuous variables). In this case, coarser-level variables correspond to higher-level, hidden descriptors of the discrete “objects” captured at finer scales. We have developed a modeling methodology and are using image recognition tasks (not just recognizing objects but also *configurations* of objects) as the initial target application.
- c) Our research in the last two years has led to major advances along a path of research adapting ideas found in fields such as compressed sensing to problems of learning models with particular “sparse” structure. In particular, we have produced a continuing stream of publications on the problem of building models for complex, high-dimensional data that expose a relatively small set of “hidden” variables which have the property that, when conditioned on these variables, the statistical structure of the original high-dimensional data is well captured by a sparse graphical model. For the Gaussian case this corresponds to extracting a decomposition of the information matrix (inverse of the covariance) of the full high-dimensional data as the sum of a sparse and a low-rank covariance matrix. Using optimization criteria that favor sparsity and small rank, we have now developed a set of theoretical guarantees and algorithms (based on semi-definite programming). As an aside, we note that this work makes contact with and, at the same time, is complementary to the direction of research described in Section 2.1(b). In particular, the decompositions here produce models that do not necessarily have tree-like structure across scales (since we do not put that constraint on the connections between hidden and original variables), and the models produced using the ideas summarized in this paragraph produce sparse inverse covariances when conditioned on the hidden variables as opposed to sparse covariances when conditioned on other scales. In addition, we have begun to extend these ideas to other related problems, including discrete-valued fields as

well as to graph decomposition in which one decomposes adjacency matrices of complex graphs into sums of far simpler ones.

- d) We have taken our work on discovering sparse structure through convex optimization considerably farther during the last year of this project. In particular, we now have a general picture of the role of convex optimization in sparse linear inverse problems, as well as a theoretical framework for graph decomposition and discovery based on convex graph invariants. In addition, we have developed significant new results for a long-standing problem in statistics, namely the decomposition of a covariance matrix into the sum of a diagonal matrix and a low-rank matrix, and we have extended these results to a new framework for learning tree-based graphical models when we are only given the statistics at the leaves of the tree. This last piece of research opens up significant areas for extension which we hope to explore in the future.
- e) One of the important areas of application of efficient graphical inference algorithms is multisensor, multitarget data association and tracking. During this past year we have continued to investigate a new graphical model representation for problems of this type that leads to algorithms that are radically different from any previously developed or used in operational systems. These algorithms, which involve real-time *smoothing* of target trajectories in order to enhance data associations offer a number of significant potential advantages. One of these is the fact that this representation makes the problem of incorporating late data – a common issue in real multi-platform surveillance applications – is a seamless operation with no additional algorithmic overhead or approximation. In addition, our experiments indicate that complexity of our algorithms scale exceptionally well with the length of the tracking interval – a dramatic difference relative to state-of-the-art algorithms. Indeed, this advantage allows the maintenance of very long tracking intervals, which allows so-called *track-stitching*, i.e., connecting track fragments separated by substantial time gaps, possible with gaps far greater than are currently feasible. This is of considerable importance in a number of operational situations of current interest, including those that are aimed at forensic analysis, e.g., to identify starting and ending locations of particular tracks that may be obscured during the middle of the tracking interval. We have now completed a first set of papers on this topic and are pursuing extensions to more complete and complex tracking contexts.
- f) We have completed a theoretical development and a paper describing methods and analyzing their performance for problems of learning sparse graphs, especially when they are designed explicitly for discrimination tasks, namely the learning of sparse graphical models for different hypotheses that, when used to form likelihood ratios, minimize resulting error probabilities when discriminating among these hypotheses. During this past year we have focused most of our attention on theoretical issues, namely analyzing the probability that methods for learning tree models make errors (i.e., learn the incorrect tree). These results, using information geometry, also provide insights into tree structures that are easier and more difficult to learn. This topic clearly overlaps strongly with the research in Section 2.3 (see brief discussion therein).
- g) We have made substantial progress in a new approach to building hierarchical graphical models in which there are potentially several layers of hidden nodes. This work has involved both an application driven part, namely the learning of hierarchical context models for the recognition of objects and groups of objects in complex scenes, and a theoretical part. In the latter we are completing a paper that provides precise results on

consistent learning of such hidden, hierarchical trees and have extended these to results on consistent estimation of hidden structure using estimated statistics. These represent significant advances which have important implications for exploitation of image-based data.

- h) We have also had a set of advances in developing information-theoretic results and guarantees on learning of tree and forest distributions from sample data. Of significant importance here is the development of scaling laws for the high-dimensional case. In addition, these results provided the theoretical foundation for some of the consistency and error analysis associated with the methods mentioned in 2.1(f). More recently we have developed new results on consistency and scaling laws for learning Ising models on general graphs.
- i) We have also begun to look at problems of performance of distributed fusion in sensor networks when the sensors are randomly located in a surveillance region. Key issues here involve how the correlation structure in both the signals and noise sensed by these distributed elements relates to the random placement structure of the sensors. For problems such as signal detection we also examine communication energy requirements associated with collecting sensor information at a fusion center. Several papers are in progress.
- j) We have completed documentation of our work on the emerging class of algorithms based on Lagrangian relaxation for MAP estimation. In this approach an overall graphical model is decomposed into a set of models each on a tractable subgraph of the original graph. Inference is then performed subject to the constraint that the estimates produced on all of these subgraphs agree. Adjoining these equality constraints via Lagrange multipliers leads to iterative algorithms in which estimates are computed on all graphs followed by modifying the decomposition to drive the estimates toward equality. For Gaussian models, in addition to guarantees of convergence for estimates, this approach also yields upper bounds on error variances which can be further tightened by optimization of the weighting used in the decomposition. Moreover, for Gaussian models we have begun to develop a framework for multiscale Lagrangian relaxation that has shown great promise for considerable speed-ups in convergence. For discrete models (e.g., as arise in problems such as data association) we have developed methods using ideas from statistical physics by replacing the maximization operation (for the computation of MAP estimates) with a temperature-dependent potential function that, when “cooled” converges to the max operator. Using this, together with adaptive methods for iteratively augmenting the graph decomposition by identifying parts of the graph in which estimates are frustrated or in competition, we have demonstrated that we can often remove duality gaps completely, yielding overall optimal solutions.
- k) We completed a body of work on the building of thinned and thus more tractable graphical models that accurately approximate the statistics of more complex models. Specifically, if we attempt to build graphical models with maximum entropy whose statistics *exactly match* those of a specified graphical model, we will, in general obtain complex models. However, if we *relax* the constraints—i.e., if we only require that the statistics of our simpler model be close to those of the more complex one—the resulting max-entropy model is frequently dramatically simpler. We have demonstrated the model-thinning power of this approach and we are now working on the problem of *adding* hidden variables in ways in which we can then perform thinning on this expanded

model. This is of particular importance in the context of multiresolution modeling (see the next topic).

- l) We have also completed our research on what we refer to as low-rank variance estimation methods for complex graphical models. The idea behind this approach is to construct low-rank approximations to the identity matrix with particular properties. Specifically, such a representation leads directly to an estimate of the variance at every node in the graph corrupted by “interference” from the cross-correlation between pairs of nodes and the dot product of the corresponding rows in the low-rank approximation to the identity. This leads to the idea of choosing the approximation to have orthogonal rows when cross-correlations are large but not worrying about their non-orthogonality if the corresponding cross-correlation is negligible. This leads to interesting graph-coloring algorithms for designing these overcomplete sets of rows, and, together with randomized choices of signs on these rows, we obtain unbiased estimates of the exact variances with guaranteed accuracy for processes with exponentially decaying correlations. For processes with long-distance correlations a variation on this approach using wavelets – and what we refer to as *spliced wavelet bases* – yields equally powerful methods for an even richer class of processes. Extension to problems involving the fusion of multiresolution data is a promising direction for the future.
- m) We have also made considerable progress on two prototypical and very important discrete optimization problems specified on graphical models, namely the so-called maximum independent set and matching problems. Such problems arise in a variety of applications including many involving resource management, sensor network organization, and optimization. Such problems are naturally cast as integer programming problems which are NP-hard. Relaxed versions of these problems can be formulated in terms of linear programs. Such a formulation can lead to integrality gaps and thus fail to give optimal answers; however in some cases the LP does indeed yield optimal solutions. Alternatively these problems can be formulated as MAP estimation problems on graphical models for which the so-called max-product algorithm provides a general purpose algorithm that is only guaranteed to yield optimal answers for graphs without loops but often works well in other contexts. We have now succeeded in providing a detailed characterization of the relationship between LP and max-product approaches. Moreover, this approach provides a very effective method for resource management in distributed fusion networks and thus makes important contact with the research in Section 2.3.
- n) We have also completed documentation of an investigation that brings together the field of decentralized team decision-making and message passing algorithms on graphs. In particular, for the case of a directed set of sensing, decision, and communication nodes (so that each node receives its own measurements together with bits from its “parent” nodes and then makes decisions resulting in bits transmitted to its “children”) we have shown that so-called person-by-person team optimization can be achieved via a message passing algorithm. This emphasizes that in communication-limited contexts with distributed agents, the agents must *organize* themselves and, in particular, design communication *protocols* for the generation and interpretation of messages within the agent network. We have now written an extensive paper on this work and demonstrated its value in designing decision networks that may differ in structure from that of the underlying variables being estimated. Moreover we have also begun to develop an

undirected version of this framework – a nontrivial extension as such a framework in principle allows feedback so that making a decision on what to communicate must also be based on the impact that that communication will have on what will be communicated *back* to the transmitting node. As with the preceding paragraph, this work involves a blend of graphical models and optimal resource utilization (in this case limited communication capacity) and hence makes contact with the research in Section 2.3.

## 2.2 Advanced Statistical Methods for Extraction and Recognition of Objects, Their Features and Geometry

The research described in this section and reported in detail in [2, 3, 6, 19, 31, 33-34, 46, 47, 63-66] has as its general objective the development of statistically robust methods for segmentation, shape estimation, and object recognition. Much of our first work in this area has focused on so-called curve evolution methods and, in particular, on developing statistically based curve evolution algorithms. However, more recently we have had successes in research directions that exploit ideas from graphical models described in the preceding subsection:

- a) The major focus of our most recent research has been on the development of hierarchical graphical models for the recognition of objects *in context* as well as the detection of objects that are out of context. Here, context refers to the learned hierarchical structure that captures the nature of scenes and the fact that certain sets of objects often occur together: cars and roads, desks and computer monitors, etc., and other objects generally don't appear together – e.g., roads and bathtubs. Using methods described in Section 2.1 for the learning of graphical models with hidden hierarchical structure, we have developed new scene-based object recognition methods that naturally exploit the dual facts that detection of particular objects may suggest particular scenes or contexts, while knowing the context may allow the detection of one type of object (e.g., a desk) to inform and enhance detection of another object (e.g., a computer mouse).
- b) The earlier component of our work in this part of our agenda has been on using curve evolution as a central component in learning decision statistics and rules from expert-labeled data. The general premise here is to design decision boundaries based on maximizing the margin – i.e., the distance to the decision boundary – of all labeled data. As the distance from a curve (or surface in higher dimensions) is directly encoded in a particular level-set function for that curve, namely the signed distance function, we are led naturally to an optimization formulation in which a margin-based cost, such as hinge loss, can be expressed directly as a function of the signed distance function from the desired decision boundary “curve.” Including a regularization term (e.g., total curve length) then leads directly to a curve evolution-based method for designing decision rules. We have demonstrated the efficacy of this approach on numerous standard data sets and also have developed theoretical results guaranteeing the consistency of the resulting estimates. In addition, we have shown how these methods can be combined with dimensionality reduction ideas in which high-dimensional data are first projected onto a lower-dimensional subspace on which the decision boundary is then determined. This area of research has obvious overlaps with that which is the focus of the third thrust of our research (see Section 2.3).
- c) We completed our work on Monte Carlo methods to sample from curve/shape distributions directly—i.e., to generate “particles” that correspond to complete curves. We have now developed a methodology for doing this – a nontrivial development as the use of Metropolis-Hastings algorithms required developing so-called *detailed balance* acceptance rules that are needed to guarantee that samples are generated by the desired shape distribution. We have also developed methods for displaying the uncertainty in the resulting extracted shapes – a feature that we believe will be of

great importance in object recognition applications. One of the appealing aspects of this sampling framework is that, with the detailed balance issue now solved, it is relatively easy to include features in the distribution that are easily used for acceptance-rejection of samples but are not easily incorporated into curve evolution methods. For example, we have demonstrated how human expert input – e.g., identifying small regions that are inside, outside, or on the boundary of the region of interest – can be easily included. Moreover, we have used ideas of graphical models to develop novel sampling methods for “2.5-dimensional object segmentation,” in which 3-D data sets (e.g., from LADAR) are segmented slice by slice, but with statistical consistency across slices accounted for via a graphical model. Several papers are in progress.

- d) One area of our most recent research is in incorporating prior information about shape into curve evolutions. This is particularly important for problems in which image SNR is low or in which the objects of interest are partially occluded. Major issues here include the development of methods for constructing prior probability distributions on shapes from examples and the incorporation of these priors into curve evolution formalisms. Our initial work in this area used a set of training examples to construct a set of “eigenshapes,” which then are used to provide a *linear* parameterization of a set of shapes, where the parameters of that linear parameterization is then estimated as part of the curve evolution process. Results on both military and medical images in both 2-D and 3-D have demonstrated that this methodology has a great deal of promise. In addition, we have been working to move beyond these linearly-parameterized methods in several different directions. The first of these methods involves postulating that the model to be learned from training examples is a mixture of two or distributions each of which is well characterized by principal component analysis. This introduces a hidden variable for each training sample—i.e., the component of the mixture to which it corresponds—which in turn leads to a new EM-based algorithm. Results demonstrate the power of this extension to classify shapes and model their variability. A second approach we are taking is that of learning nonparametric models for shapes given a set of training samples. Nonparametric density estimation methods require the use of a distance metric between pairs of shapes, and our work has led us to use two natural metrics, each of which leads to a different curve evolution. Both of these have been shown to have considerable promise for recognizing and segmenting shapes that can have considerable variability or be subject to partial occlusion. We are also developing new methods that can incorporate human or expert input – e.g., in the form of partial segmentations – to help guide both curve evolution as well as Monte Carlo sampling.

### **2.3 Machine Learning and Optimization Methods for Robust Fusion, and Effective Use of Limited and Distributed Resources**

The research described in this section deals with methods for complex signal, image, and data analysis using methods of machine learning and optimization- based formulations. Our research is described in [16, 18, 20-21, 24-25, 29-34, 36-39, 42-44, 46-48, 52-54, 59-62]. Our research has led to the following lines of inquiry and results:

- a) We have had major successes and considerable publicity for our work on using Hierarchical Dirichlet Processes (HDP) in learning target motion patterns and, more generally, multiple modes of dynamic behavior for complex systems represents a major new thrust for our research on learning models for complex dynamic phenomena. In particular, we have developed new hidden Markov model (HMM) and switching state space models that do not presuppose any knowledge of the number of modes to be captured in these switching models, the transition probabilities between these modes, or the dynamic behavior for each mode. Our work to date has shown considerable promise, including demonstrations on extracting models of the complex behavior of “bee dances” (in which bees engage in complex motion patterns to signal the location of a food source; a problem that is an obvious surrogate for patterns of interest in military scenarios), on detecting major economic events from the dynamic behavior of stock indices, and the extraction and segmentation of audio signals in which an unknown number of unknown speakers are engaged in conversation (where we do not know what any speaker sounds like nor do we know when each is speaking). All of these results are being documented in a series of papers. We have also initiated extensions to allow semi-Markov processes and also a very powerful extension involving extracting modes of behavior that are exhibited by groups of objects (in which each object may exhibit only a subset of these modes).
- b) During the past two years we have developed new methods that go beyond those described in Section 2.3(a) above. In particular, the restriction to hidden *Markov* behavior in our earlier work, while significant, has limitations in terms of expressivity in terms of capturing memory in complex data. Motivated by this observation, we have developed an extension of our HDP framework to hidden *semi-Markov models* (*HSMMs*). Such models separate the designation of different modes of behavior from the detailed definition of system *state* and lead to very powerful new models with considerably greater expressivity (e.g., Morse code dots and dashes are difficult to represent with HMMs without many states, while they are very easily described with HSMMs). Moreover and very interestingly, the extension to HSMMs suggests much more efficient methods for inference and sampling for HDP-HMM models as described in 2.3(a).
- c) In Section 2.1(f) we described one of the directions of research that lies at the intersection of graphical models and learning, namely the problem of learning tractable graphical models from data, where the criterion used is not model accuracy but *model utility* – in particular in hypothesis testing/classification applications in which the challenge is discriminating between two high-dimensional probability distributions given a limited set of training data. As one would expect, if vast amounts of data are available, the models learned for the two different probability distributions revert to the best models for each individually. However, when data are limited, the results can be significantly different – i.e., from these limited data what we really desire are models that highlight *saliency*, the significant differences between hypotheses. In particular, we have now developed very efficient models for building discriminative tree and forest models from sample data in order to optimize discrimination performance as measured by so-called J-divergence. Very

importantly, the algorithm for the optimal solution to this problem is greedy, so that it starts by incorporating the most salient difference between the observed data features under the different hypotheses and then successively adds additional features if they add to discrimination performance. This is of potentially great value in many contexts in which high-dimensional data need to be processed but sufficient data are not available to build accurate models (or building such models is computationally intractable). Applications ranging from hyperspectral data analysis to multimodal fusion for object classification will benefit from this line of research. In addition, we have shown how we can use boosting to build discriminators that use a collection of tree likelihood functions (and hence function in a manner very similar to that of models on more complex graphs than trees). In addition, we have very recently made significant theoretical progress in providing precise results that make it clear that focusing on saliency can greatly reduce the number of training samples needed for discriminative learning, a fact that is extremely important in applications such as automatic target recognition.

- d) A continuing and very active component of our research focuses on so-called sparsity-based signal and image processing. On the theoretical side, we have recently documented significant new results on so-called compressed sensing, a topic of great current interest in research and practice in which signals that are known to be sparse in a particular basis (i.e., have a relatively small number of nonzero coefficients) can be faithfully reconstructed from surprisingly small sets of measurements (as long as those measurements are “diffuse” with respect to the basis in which the signal to be recovered is sparse). In our work we have shown that if one solves this problem recursively, adding data samples at each step, one can not only develop very precise and simple stopping rules, but when one stops, in general even fewer data points are required. In addition, as mentioned in Section 2.1(e), we have adapted some of the ideas behind compressed sensing – namely variational formulations employing regularized norms such as  $l_1$  – that are used as surrogates for sparsity. In our case, we have used both  $l_1$  to prefer sparsity in learned graphical models as well as the so-called nuclear norm (sum of singular values), a surrogate for rank to learn hidden models for complex data, in which the low-rank portion corresponds to the influence of a set of hidden variables, and the sparse portion corresponds to the conditional graphical structure of the observed variables when conditioned on the hidden variables. As mentioned in Section 2.1(c) we have obtained theoretical results and developed algorithms to find such decompositions which provide very attractive models for inference for Gaussian processes.
- e) We have developed a set of results on constructing or learning decision rules for complex data. One part of this work deals with the problem of modeling experts in terms of their prior models for a set of hypotheses. Using a well-documented phenomenon that humans tend to categorize items, we have developed an approach to optimal quantization of prior probabilities in hypothesis testing problems. This leads to nontrivial and important insights into how such categorization can bias decision-making. Interestingly, this work then served as the launching point for work described in part in Section 2.2(b) on learning decision rules and decision regions from expert-labeled data. In Section 2.2(b) we described our work on using curve evolution methods to determine decision boundaries that maximize the margin in

decision-making. We have also developed methods aimed at dimensionality reduction, i.e., at projecting high-dimensional data onto lower-dimensional subspaces that contain the discriminating information used in these expert-labeled examples. We have shown how we can couple this either with curve evolution methods or with support vector machines and have performed theoretical analysis providing conditions for consistency and also demonstrating the value of dimensionality reduction when limited training data are available – i.e., in contexts in which reducing dimensionality can greatly reduce the tendency toward overfitting. In addition we have extended these ideas to problems in distributed fusion, in which sensors are organized into a directed fusion network and each sensor must perform dimensionality reduction before forwarding its data to subsequent nodes in the network and ultimately to the fusion center which has the objective of taking all information that reaches it and making maximum-margin decisions. Very importantly, the optimization of the different sensors’ dimensionality reduction computations involves message-passing propagating information through the fusion network.

- f) As mentioned in Section 2.1, some of our work on graphical models has led to new methods for optimizing resource utilization in distributed fusion networks. In particular, the research mentioned in Section 2.1 (k) includes new results on algorithms for problems such as optimal formation of a communication network for a set of distributed sensors, in which the cost to be optimized involves weights on each potential link trading off informational value of that link with the power required for communication using it. The research described in Section 2.1 (m) involves the development of distributed algorithms for organizing the signaling among a set of sensors once the communication network has been established. In particular, in this methodology, sensors must develop a fusion protocol so each sensor knows how to interpret information sent to it by other sensors and then knows how to process these, together with its own local data to produce signals to send to other sensors in order to optimize an overall team objective that is a weighted combination of decision error costs (where decisions are made by a subset of the sensing nodes) and total communication required. Interestingly the process of determining this fusion protocol admits a message-passing implementation itself, so that the organization of the sensor network can be accomplished in a distributed manner.
- g) Sparsity also remains an important part of our work on variational methods to produce enhanced images and reconstructions for SAR, ISAR, and more general array processing applications. In particular, by putting particular penalties (e.g.,  $L_p$ , with  $p < 1$ ) either on the reconstructed image or on the gradient of the reconstructed image, we have shown that we can produce remarkably sharp images of point scatterers or regions and can also correct for phase errors due to target motion—an extremely important problem in SAR imaging of moving targets or to other sources (including timing errors to array element location errors). Moreover, in contrast to many other superresolution methods (e.g., MUSIC, Capon’s method), our method can resolve multiple scattering effects that are highly correlated—e.g., due to the presence of multipath effects. In one part of our research we have developed new variational approaches for array processing that work well for broadband sources and, in particular, for sources that generate multiple harmonics (e.g., as are present in any

motor or machinery). In another component of our research we have taken a deeper look at marrying SAR physics with nonparametric statistical learning methods for constructing probabilistic models for multiresolution imagery. In particular consider the formation of SAR imagery based on a given full aperture of data. If we use the entire aperture, we obtain imagery at the finest resolution resolvable using that data. However, to do this we in essence must assume that all scattering is isotropic, i.e., that the response from significant scatterers is constant across the entire aperture. For many important scattering mechanisms this is not the case at all, and this anisotropy is critical to distinguishing one scatterer type from another. Suppose then, that in addition to forming an image using the entire aperture, we also form three images each using half of the aperture: one image using the right half, one the left, and one using a centered half-aperture. If indeed there are anisotropic scatterers, we might expect that there would be differences in the responses in each of these half-apertures and hence in the images formed using them (note that these images would have pixel sizes twice as large as the ones in the finest scale imagery). Iterating this process, we can imagine forming a vector of images at each of a sequence of scales corresponding to progressively smaller subapertures. By looking across scale, then, we would expect not only to find statistical variability due to speckle but also any evidence of anisotropic scattering manifesting itself in statistically significant differences in pixel intensities in images formed using different subapertures. We have initiated an effort in this area that employs the “sparseness prior” variational framework described in the preceding paragraph. Initial results provide the basis for some new “best basis” methods for imaging that avoid exhaustive search of subapertures through a modified coarse-to-fine search with intelligent back-tracking. We believe that there is much more that can be done in this area. For example, one very promising direction for future work is that of coupling these front-end algorithms with back-end object recognition using the framework of Dirichlet processes for object recognition described in the preceding section. In particular, we expect that by building object models that couple object models with anisotropy properties we will be able to develop algorithms in which object-level hypotheses will drive front-end signal processing. This offers the possibility of a significant conceptual and algorithmic leap over current methods (e.g., the current form of the so-called “PEMS Loop” in the algorithms developed under the MSTAR program).

- h) We have also developed a new, first principles probabilistic approach to Markov modeling on trees, together with a start on the nontrivial generalization to graphs with loops. Interestingly this approach identifies reduced sets of conditional independence relationships that need to be verified either in determining if a particular set of variables are Markov or in *designing* hidden variable representations to ensure Markovianity. The former interpretation of our results is of great importance in the context of the estimation of the *structure* among a set of observed variables—e.g., to identify statistical links among them as well as conditional independencies, a topic sometimes referred to as *link discovery*. This is closely related to our recently-initiated work on learning models for coordinated motion patterns of multiple objects. One long-term objective of this portion of our work is to tie it in with the Dirichlet process-based methods described in (a) in order to develop methods for automatically determining such coordinated motion models on the fly.

### **III. Personnel**

The following is a list of individuals who have worked on research supported in whole or in part by the Air Force Office of Scientific Research under Grant FA9559-08-1-0180:

Prof. Alan S. Willsky, Edwin Sibley Webster Professor of Electrical Engineering, MIT  
Dr. John Fisher, Senior Research Scientist, MIT Comp.Sci. and AI Laboratory  
Dr. Mujdat Cetin, Research Scientist, MIT Lab. For Information and Decision Systems  
Dr. Sujay Sanghavi, Post-doctoral Researcher, MIT Lab. For Info. and Decision Systems  
Dr. Justin Dauwels, Research Scientist, MIT Lab. for Information and Decision Systems  
Dr. Animashree Anandkumar, Post-Doctoral Researcher, MIT Lab. for Info. & Decision Systems  
Dr. David Wingate, Research Scientist, MIT Lab. for Information and Decision Systems  
Dr. Oliver Kosut, Post-doctoral Researcher, MIT Lab. for Information and Decision Systems  
Dr. Junmo Kim, graduate student (Ph.D. completed)  
Dr. Ayres Fan, graduate student (Ph.D. completed)  
Dr. Dmitry Malioutov, graduate student (Ph.D. completed)  
Dr. Jason Johnson, graduate student (Ph.D. completed)  
Mr. Michael Chen, graduate student (MEng completed)  
Dr. Emily Fox, graduate student (Ph.D. completed)  
Dr. Kush Varshney, graduate (Ph.D. completed)  
Dr. Venkat Chandrasekaran, graduate student (Ph.D. completed)  
Dr. Myung Jin Choi, graduate student (Ph.D. completed)  
Dr. Vincent Tan, graduate student (Ph.D. completed)  
Mr. Ying Liu, graduate student (SM completed; Ph.D. in progress)  
Mr. Matthew Johnson, graduate student (SM completed; Ph.D. in progress)  
Mr. James Saunderson, graduate student (SM completed; Ph.D. in progress)

## IV. Publications

The publications listed below represent papers, reports, and theses supported in whole or in part by the Air Force Office of Scientific Research under Grant FA9559-08-1-0180:

- [1] A. Ihler, J.W. Fisher III, and A.S. Willsky, "Nonparametric Dynamical Modeling with Applications to Signature Authentication," in preparation for submission to *IEEE Trans. on Signal Processing*.
- [2] A. J. Kim, J. W. Fisher III, and A. S. Willsky, "Multiresolution/Multiaperture Algorithms for Characterizing and Imaging Anisotropy in SAR Imagery," in preparation.
- [3] K.R. Varshney, M. Cetin, J.W. Fisher, III, and A.S. Willsky, "Sparse Representation in Structured Dictionaries with Application to Synthetic Aperture Radar," *IEEE Trans. on Signal Processing*, Vol. 56, No. 8, August 2008, pp. 3548-3561.
- [4] E.B. Sudderth, A.T. Ihler, M. Isard, W.T. Freeman, and A.S. Willsky, "Nonparametric Belief Propagation," invited paper for *Communications of the ACM*
- [5] J. Johnson, "Information-geometric Methods for Approximate Graphical Modeling," Ph.D. thesis, August 2008.
- [6] W. Sun, M. Cetin, R. Chan, V. Reddy, G. Holmvang, V. Chandar, and A.S. Willsky, "Learning the Dynamics and Time-Recursive Boundary Detection of Deformable Objects," *IEEE Trans. on Image Processing*, Vol. 17, No. 11, Nov. 2008, pp. 2186-2200.
- [7] E.B. Sudderth, M. Mandel, W.T. Freeman, and A.S. Willsky, "Visual Hand Tracking Using Occlusion Compensated Message Passing," submitted to *Computer Vision and Image Understanding*.
- [8] D.M. Malioutov, "Inference for Gaussian graphical models: Analysis and algorithms," Ph.D. thesis, May 2008.
- [9] M.J. Choi, V. Chandrasekaran, D.M. Malioutov, J.K. Johnson, and A.S. Willsky, "Multiscale Stochastic Modeling for Tractable Inference and Data Assimilation," invited paper, *Computer Methods in Applied Mechanics and Engineering* on stochastic modeling for multiscale/multiphysics problems, Vol. 197, Aug. 2008, pp. 3492-3515.
- [10] V. Chandrasekaran, J.K. Johnson, and A.S. Willsky, "Estimation in Gaussian Graphical Models Using Tractable Subgraphs: A Walk-Sum Analysis," *IEEE Trans. on Signal Processing*, Vol 56, No.5, May 2008, pp. 1916-1930.

- [11] J.K. Johnson, V. Chandrasekaran, and A.S. Willsky, "Maximum Entropy Relaxation for Graphical Model Selection," in preparation for submission to *Journal of Machine Learning Research*.
- [12] D.M. Malioutov, J. Johnson, M.J. Choi, and A.S. Willsky, "Low -Rank Variance Approximation in GMRF Models: Single and Multi-Scale Approaches," *IEEE Trans. on Signal Processing*, Vol. 56, No. 10, Oct. 2008, pp. 4621-4634.
- [13] O.P. Kreidl and A.S. Willsky, "An Efficient Message-Passing Algorithm for Optimizing Decentralized Detection Networks," *IEEE Trans. on Automatic Control*, Vol. 55, No. 3, March 2010, pp. 563-578.
- [14] O.P. Kreidl and A.S. Willsky, "Decentralized Detection in Undirected Network Topologies," in preparation for submission to *IEEE Trans. on Signal Processing*.
- [15] M. Cetin, L. Chen, J.W. Fisher, III, A.T. Ihler, O.P. Kreidl, R.L. Moses, M.J. Wainwright, J.L. Williams, and A.S. Willsky, "Graphical Models and Fusion in Sensor Networks," Chapter to appear in the book *Wireless Sensor Networks" Signal Processing and Communication Perspectives*, A. Swami, L. Tong, and Q. Tiao, editors, John Wiley & Sons, Ltd., England; to appear.
- [16] V. Tan, J.W. Fisher, III, and A.S. Willsky, "Learning Max-Weight Discriminative Forests," ICASSP 2008.
- [17] M.J. Choi, V. Chandrasekaran, and A.S. Willsky, "Maximum Entropy Relaxation for Multiscale Graphical Model Selection," ICASSP 2008.
- [18] D. Malioutov, S. Sanghavi, and A.S. Willsky, "Compressed Sensing with Sequential Observations," ICASSP 2008.
- [19] W. Sun, M. Cetin, R. Chan, and A.S. Willsky, "Segmentation of the Evolving Left Ventricle by Learning the Dynamics," to be presented at ISBI'08, the Fifth International Symposium on Biomedical Imaging, May 2008, Paris, France.
- [20] E.B. Fox, E.B. Sudderth, M. Jordan, and A.S. Willsky, "An HDP-HMM for Systems with State Persistence," ICML 2008.
- [21] K.R. Varshney and A.S. Willsky, "Supervised Learning of Classifiers via Level Set Segmentation," Machine Learning for Signal Processing Workshop 2008, Cancun, Mexico.
- [22] S. Sanghavi, D. Malioutov, and A.S. Willsky, "Belief Propagation and LP Relaxation for Weighted Matching in General Graphs," accepted for publication in *IEEE Trans. on Information Theory*.

- [23] S. Sanghavi, D. Shah, and A.S. Willsky, "Message Passing for Maximum Weight Independent Sets, *IEEE Trans. on Information Theory*, Vol. 55, No. 11, Nov. 2009, pp. 4822-4834.
- [24] V.Y.F. Tan, S. Sanghavi, J.W. Fisher, and A.S. Willsky, "Learning Graphical Models for Hypothesis Testing and Classification," accepted for publication in the *IEEE Trans. on Signal Processing*.
- [25] E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky, "Nonparametric Bayesian Learning of Switching Dynamical Systems," NIPS 2008.
- [26] M.J. Choi, V. Chandrasekaran, and A.S. Willsky, "Gaussian Multiresolution Models: Exploiting sparse Markov and Covariance Structure," International Conference on Machine Learning, June 2009.
- [27] O.P. Kreidl and A.S. Willsky, "Decentralized Detection with Long-Distance Communication," Asilomar Conference, October 2008.
- [28] Z.M. Chen, L. Chen, M. Cetin, and A.S. Willsky, "Efficient Multi-Target Tracking Using Graphical Models," in preparation for submission to *Advances in Information Fusion*.
- [29] E.B. Fox, E.B. Sudderth, M.I. Jordan, A.S. Willsky, "Nonparametric Bayesian Identification of Jump Systems with Sparse Dependencies," Proceedings of the 15th IFAC Symposium on System Identification, Saint-Malo, France July 2009.
- [30] Kush R. Varshney and Lav R. Varshney. Minimum Mean Bayes Risk Error Quantization of Prior Probabilities. 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing, p. 3445-3448, Las Vegas, Nevada, April 2008.
- [31] Kush R. Varshney *Frugal Hypothesis Testing and Classification*, Ph.D. thesis, MIT, February 2010.
- [32] Kush R. Varshney and Lav R. Varshney. Quantization of Prior Probabilities for Hypothesis Testing. *IEEE Transactions on Signal Processing*, vol. 56, issue 10, p. 4553-4562, October 2008.
- [33] Kush R. Varshney and Alan S. Willsky. Classification Using Geometric Level Sets. *J. of Machine Learning Research*, Vol. 11, Feb. 2010, pp. 491-516.
- [34] Kush R. Varshney and Alan S. Willsky. Learning Dimensionality-Reduced Classifiers for Information Fusion. International Conference on Information Fusion, Seattle, Washington, July 2009, awarded one of the Best Student Paper Awards.
- [35] Myung Jin Choi, Venkat Chandrasekaran, and Alan S. Willsky, "Gaussian Multiresolution Models: Exploiting Sparse Markov and Covariance Structure," *IEEE Trans. on Signal Processing*, Vol. 58, March 2010, pp. 1012-1024.

- [36] E.B. Fox, E.B. Sudderth, M.I. Jordan, A.S. Willsky, "The Sticky HDP-HMM: Nonparametric Bayesian Hidden Markov Models with Persistent States," accepted for publication in *Annals of Applied Statistics*.
- [37] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Sparse and Low-Rank Matrix Decompositions," 15th IFAC Symposium on System Identification (SYSID 2009), Saint-Malo, France, July 2009.
- [38] Vincent Y. F. Tan, Animashree Anandkumar, Lang Tong and Alan Willsky, "A Large-Deviation Analysis for the Maximum-Likelihood Learning of Tree Structures," 2009 IEEE International Symposium on Information Theory (ISIT09), Seoul, Korea, Jun 28 - Jul 3, 2009.
- [39] O.P. Kreidl and A.S. Willsky, "Decentralized detection with long-distance communication," Proceedings of the 42nd Annual Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, Oct 2008.
- [40] A. Anandkumar, J.E. Yukich, L. Tong, A. Willsky, "Detection Error Exponent for Spatially Dependent Samples in Random Networks," Proc. ISIT, Seoul, South Korea, July 2009.
- [41] Zhexu (Michael) Chen, Lei Chen, Mijdat Cetin, and Alan S. Willsky, "An Efficient Message Passing Algorithm for Multi-Target Tracking," 2009 International Conference on Information Fusion.
- [42] Emily B. Fox, "Bayesian Nonparametric Learning of Complex Dynamical Phenomena," Ph.D. thesis, MIT, July 2009.
- [43] S. Sanghavi, D. Malioutov, and A.S. Willsky, "Networking Sensors via Belief Propagation," Allerton Conference. 2008.
- [44] D. Malioutov, S. Sanghavi, and A.S. Willsky, "Sequential Compressed Sensing," special issue of the *IEEE Journal on Selected Topics in Signal Processing* on Compressive Sensing, Vol. 4, No. 2, April 2010, pp. 435-444.
- [45] Y. Liu, "Inference in Gaussian Graphical Models: Algorithms Exploiting Self-Avoiding Walks and Cycle Bases," SM thesis, MIT, May 2010.
- [46] K.R. Varshney and A.S. Willsky, "Linear Dimensionality Reduction for Margin-Based Classification: High-Dimensional Data and Sensor Networks," accepted for publication in *IEEE Trans. on Signal Processing*.
- [47] K.R. Varshney and A.S. Willsky, "Classification Using Geometric Level Sets," *J. of Machine Learning Research*, Vol. 11, Feb. 2010, pp. 491-516.

- [48] E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky, "Sharing Features Among Dynamical Systems with Beta Processes," *NIPS* 2009.
- [49] V.Y.F. Tan, A. Anandkumar, and A.S. Willsky, "Learning Gaussian Tree Models: Analysis of Error Exponents and Extremal Structures," *IEEE Trans. on Signal Processing*, Vol. 58, No. 5, May 2010, pp. 2701-2714.
- [50] V.Y.F. Tan, A. Anandkumar, L. Tong, and A.S. Willsky, "A Large-Deviation Analysis for the Maximum-Likelihood Learning of Markov Tree Structures," accepted for publication in *IEEE Trans. on Information Theory*.
- [51] V.Y.F. Tan, A. Anandkumar, and A.S. Willsky, "How do the Structure and Parameters of Tree Gaussian Models Affect Structure Learning," Allerton Conf., Sept. 2009.
- [52] V.Y.F. Tan and A.S. Willsky, "Necessary and Sufficient Conditions for High-Dimensional Salient Subset Recovery," IEEE Int'l. Symposium on Information Theory (ISIT10).
- [53] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A.S. Willsky, "Rank-Sparsity Incoherence for Matrix Decomposition," accepted for publication in *SIAM J. on Optimization*.
- [54] V. Chandrasekaran, "Algebraic and Spectral Approaches to Statistical Modeling and Graphs," Ph.D. thesis, May 2011.
- [55] M.J. Choi, V. Chandrasekaran, and A.S. Willsky, "Exploiting Sparse Markov *and* Covariance Structure in Multiresolution Models," Proc. Int'l. Conf. on Machine Learning (ICML), Montreal, Canada, June 2009.
- [56] M.J. Choi, "Hierarchical Graphical Models with Latent Variables," Ph.D. thesis, Jan. 2011.
- [57] M.J. Choi, J.J. Lim, A. Torralba, and A.S. Willsky, "Exploiting Hierarchical Context on a Large Database of Object Categories," submitted to IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2010.
- [58] Y. Liu, V. Chandrasekaran, A. Anandkumar, and A.S. Willsky, "Feedback Message Passing for Inference in Gaussian Graphical Models," IEEE Int'l. Symposium on Information Theory (ISIT10).
- [59] M. Johnson and A.S. Willsky, "The Hierarchical Dirichlet Process Hidden Semi-Markov Model," poster presentation at UAI 2010.
- [60] M. Johnson, "The Hierarchical Dirichlet Process Hidden Semi-Markov Model," SM thesis, May 2010.

- [61] E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky, "Bayesian Nonparametric Methods for Learning Markov Switching Processes," special issue of the *IEEE Signal Processing Magazine* on graphical models in signal processing, Vol. 27, No. 6, Nov. 2010, pp. 43-54.
- [62] E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky, "Bayesian Nonparametric Inference of Switching Dynamic Linear Models," *IEEE Trans. on Signal Processing*, Vol. 59, No. 4, April 2011, pp. 1569-1585.
- [63] M.J. Choi, J.L. Lim, A. Torralba, and A.S. Willsky, "Exploiting Hierarchical Content on a Large Database of Object Categories," IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) San Francisco, June 2011.
- [64] M.J. Choi, V.Y.F. Tan, A. Anandkumar, and A.S. Willsky, "Learning Latent Tree Graphical Models," accepted for publication in the *Journal of Machine Learning Research*.
- [65] M.J. Choi, V.Y.F. Tan, A. Anandkumar, and A.S. Willsky, "Consistent and Efficient Reconstruction of Latent Tree Models," Allerton Conference 2010.
- [66] M.J. Choi, J.L. Lim, A. Torralba, and A.S. Willsky, "A Tree-Based Context Model for Object Recognition," submitted to *IEEE Trans. on Pattern Analysis and Machine Intelligence*.
- [67] V.Y.F. Tan, M. Johnson, and A.S. Willsky, "Necessary and Sufficient Conditions for High-Dimensional Salient Subset Recovery," IEEE Int'l. Symposium on Information Theory (ISIT10).
- [68] V.Y.F. Tan, A. Anandkumar, and A.S. Willsky, "Learning High-Dimensional Markov Forest Distributions: Analysis of Error Rates," accepted for publication in the *Journal of Machine Learning Research*.
- [69] V.Y.F. Tan, A. Anandkumar, and A.S. Willsky, "Scaling Laws for Learning High-Dimensional Markov Forests," Allerton Conference 2010.
- [70] A. Anandkumar, V.Y.F. Tan, and A.S. Willsky, "High-Dimensional Structure Reconstruction of Ising Models on Sparse Random Graphs," Allerton Conference 2010.
- [71] V. Chandrasekaran, B. Recht, P.A. Parrilo, and A.S. Willsky, "The Convex Algebraic Geometry of Linear Inverse Problems," Allerton Conference 2010.
- [72] V. Chandrasekaran, P.A. Parrilo, and A.S. Willsky, "Latent Variable Graphical Model Selection via Convex Optimization," invited paper for Allerton Conference 2010.
- [73] A. Anandkumar, J.E. Yukich, and A.S. Willsky, "Scaling Laws for Random Spatial Graphical Models," IEEE Int'l. Symposium on Information Theory (ISIT10).

- [74] A. Anandkumar, J.E. Yukich, and A.S. Willsky, “Random Spatial Graphical Models,” to be submitted to *Random Structures and Algorithms*.
- [75] A. Anandkumar, B. Bolobas, P. Balister, and A.S. Willsky, “Energy-Latency Tradeoff for In-Network Function Computation in Random Networks,” accepted for INFOCOM 2011.
- [76] V.Y.F. Tan and A.S. Willsky, “Sample Complexity for Topology Estimation in Networks of LTI Systems,” accepted for presentation at the IFAC World Congress, 2011.
- [77] A. Anandkumar, V.Y.F. Tan, and A.S. Willsky, “High-Dimensional Robust Structure Learning of Ising Models on Sparse Random Graphs,” submitted to *Annals of Statistics*.
- [78] V. Chandrasekaran, B. Recht, P. Parrilo, and A.S. Willsky. “The Convex Geometry of Linear Inverse Problems,” submitted to *Foundations of Computational Mathematics*.
- [79] V. Chandrasekaran, P. Parrilo, and A.S. Willsky. “Convex Graph Invariants,” submitted to *SIAM Review*.
- [80] J. Saunderson, V. Chandrasekaran, P. Parrilo, and A.S. Willsky, “Tree-Structured Statistical Modeling via Convex Optimization,” submitted to the 50<sup>th</sup> IEEE Conf. on Decision and Control, Dec. 2011.
- [81] J. Saunderson, “Convex Optimization Methods for Covariance Decomposition and Tree-Structured Modeling,” SM thesis, May 2011.

## V. INTERACTIONS/TRANSITIONS

In this section we summarize our and plans for transitions associated with research supported by AFOSR Grant FA9559-08-1-0180, as well as listing some important honors received by members of our research team.

### Honors and Recognition

- (1) Dr. Junmo Kim, Dr. Mujdat Cetin, and Prof. Alan Willsky were awarded the 2008 Best Paper Award for their paper “Nonparametric Shape Priors for Active Contour-Based Image Segmentation,” in the journal *Signal Processing*.
- (2) Prof. Alan S. Willsky was appointed Director of MIT’s Laboratory for Information and Decision Systems
- (3) The research of Dr. Emily Fox was chosen by AFOSR for a research highlight and has also been featured in *Signal* magazine.
- (4) Dr. Kush Varshney received a Best Student Paper Award for his paper at the 2009 International Conference on Information Fusion.
- (5) Prof. Alan S. Willsky was awarded the 2010 IEEE Signal Processing Society Technical Achievement Award.
- (6) Dr. Fox received the Jin-Au Kong Outstanding Doctoral Thesis Prize from MIT’s Dept. of EECS.
- (7) Dr. Fox received the Savage Award for the best Ph.D. thesis in Applied Methodology in Bayesian Statistics.
- (8) Prof. Willsky was elected to the National Academy of Engineering in 2010.
- (9) Dr. Dmitry Malioutov, Dr. Mujdat Cetin, and Prof. Willsky received the 2010 IEEE Signal Processing Society Best Paper Award for their paper “A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays,” in the *IEEE Trans. on Signal Processing*.

### Participation/Presentation at Meetings

In addition to the a number of invited and contributed talks presented at various meetings, we also make note of the following:

- (1) Prof. Willsky and many of the students, scientists, and post-docs in his group have given a continuing series of lectures on their research at MIT Lincoln Laboratory.
- (2) Prof. Willsky was the only academic participant at the 2010 meeting on Mission-Focused Autonomy held at JIATF-S in Key West Florida in June 2010.
- (3) Prof. Willsky gave a plenary address at the 2010 Machine Learning Workshop associated with the Neural Information Processing Systems Symposium.

### Consultative and Advisory Functions

We continue to be actively engaged in a number of activities relevant to the research being performed under our AFOSR grant:

- (1) Prof. Willsky has regularly acted as a consultant to BAE Systems Advanced Information Technologies (BAE-AIT; formerly Alphatech, Inc.) in a number of

- research projects including ones that represent direct transitions of the technology being developed under our AFOSR Grant.
- (2) Prof. Willsky served on the Senior Review Panel for DARPA's POSSE (Persistent, Operational Surface Surveillance and Engagement) Program which is aimed at rapid deployment of advanced ISR systems to active areas of conflict (note that all of the other members of the panel are either retired 3- and 4-star generals or individuals who previously served as Deputy Assistant Secretaries of Defense).
  - (3) Prof. Willsky has recently initiated consulting activities with Parietal Systems, Inc. and is actively involved in transitions of his research to programs being conducted and envisioned.

## **Transitions**

The following represent some of the ongoing transitions of our work as well as some plans for future transitions:

- (1) Our work on Lagrangian Relaxation Methods has been incorporated into BAE System Advanced Information Technologies (BAE-AIT) All-Source Track and ID Fusion (ATIF) System.
- (2) Our work on sensor resource management has been transitioned to Lincoln Laboratory.
- (3) Dr. Mujdat Cetin's methods for sparse regularization for radar signal processing and SAR analysis have been transitioned to AFRL/SN, and Dr. Cetin, in collaboration with Prof. Randy Moses of Ohio State University have been working toward enhancing this transition.
- (4) We are moving forward with engineers at Parietal Systems for the transition of our new graphical-model-based approach to multi-sensor, multi-target tracking. The
- (5) We are actively pursuing at BAE-AIT, Lincoln Laboratory, and Parietal Systems, Inc. on transitioning our methods for automatic learning of behavior models for targets and other dynamically evolving phenomena using the emerging class of models based on Dirichlet Processes. In particular Parietal Systems is working on several Air Force SBIR programs that aim explicitly at that transition.
- (6) Our work on machine-learning-based methods for multisensor fusion has been transitioned to BAE-AIT where it has been applied to problems of audio-video fusion.